

Large Deviations of the Smallest Eigenvalue of the Wishart-Laguerre Ensemble

Eytan Katzav and Isaac Pérez Castillo

Department of Mathematics, King's College London, Strand, London WC2R 2LS, United Kingdom

We consider the large deviations of the smallest eigenvalue of the Wishart-Laguerre Ensemble. Using the Coulomb gas picture we obtain rate functions for the large fluctuations to the left and the right of the hard edge. Our findings are compared with known exact results for $\beta = 1$ finding good agreement. We also consider the case of almost square matrices finding new universal rate functions describing large fluctuations.

PACS numbers: 05.40.-a, 02.10.Yn, 02.50.Sk, 24.60.-k

In the early 50s, when not much was known about the intricacies of complex atomic nuclei, Wigner suggested to replace the underlying physics of the problem by its apparent statistical features [1]. Surprisingly, it turned out that such statistical approach was more useful than anyone could have anticipated, enabling him to work out the nuclei level spacing distribution. Random Matrix Theory (RMT) has played a central role in various branches of science, since its first appearance in Statistics by Wishart in 1928 [2], through QCD [3], random graphs [4], wireless communications [5] and computational biology [6] - to mention a few. And although through more than half a century different, seemingly unrelated, problems have been linked via RMT, certain questions about eigenvalue distributions are still poorly explored. One such question is the distribution of the smallest eigenvalue.

In physics, for instance, classical disordered systems offer the ideal context where RMT concepts and tools may be applied. Here, complicated systems are simplified by using a random Hamiltonian, where the smallest eigenvalue is of special interest as it is associated with the ground state. In quantum entanglement, the smallest eigenvalue is a useful measure of entanglement [7].

In mathematics, the minimal eigenvalue appears naturally in many contexts, such as the study of the geometry of random polytopes [8]. It is also extremely relevant for the question of invertibility of random matrices [9], and recently played an important role in the exploding field of compressive sensing [10], where fluctuations of the minimal eigenvalue set the bounds on the number of random measurements needed to fully recover a sparse signal.

In statistics, a very important technique used to detect hidden patterns in complex, high-dimensional datasets is called "Principal Components Analysis" (PCA). The idea is to take a data matrix \mathbf{X} , and to transform its covariance matrix $\mathbf{W} = \mathbf{X}^\dagger \mathbf{X}$ into a new coordinate system such that the greatest variance by any projection of the data comes to lie on the first coordinate. Technically speaking, one identifies eigenvalues and eigenvectors of \mathbf{W} , and ignores the components corresponding to the lowest eigenvalues, as these eigenmodes contain the least important information. The smallest eigenvalue of \mathbf{W} determines the largest eigenvalue of \mathbf{W}^{-1} , and is important

in Hotelling's T-square distribution [11], for example. The purpose of this work is to provide a simple physical method, based on the Coulomb gas method in statistical physics [12], that allows us to compute analytically the probability of the smallest eigenvalue in the Wishart-Laguerre ensemble.

We consider an ensemble $G(M, N)$ of $M \times N$ rectangular matrices $\mathbf{X} \in G(M, N)$, which are drawn from a Gaussian distribution $P(\mathbf{X}) \propto \exp \left[-\frac{\beta}{2} \text{Tr}(\mathbf{X}^\dagger \mathbf{X}) \right]$, where β is the Dyson index with classical values $\beta = 1, 2$, and 4 , corresponding to the real, complex and quaternionic cases respectively. The Wishart ensemble $W(M, N)$ is defined as the set of $N \times N$ matrices $\mathbf{W} = \mathbf{X}^\dagger \mathbf{X}$. If λ denotes the N eigenvalues of a Wishart matrix, their joint PDF reads

$$P(\lambda) = \frac{1}{Z(0)} e^{-\beta F(\lambda)/2} \quad (1)$$

with $Z(0)$ a normalisation constant and $F(\lambda)$ defined as:

$$F(\lambda) = \sum_{i=1}^N \lambda_i - u \sum_{i=1}^N \ln(\lambda_i) - \sum_{i \neq j} \ln |\lambda_i - \lambda_j| \quad (2)$$

where $u = 1 + M - N - \frac{2}{\beta}$.

We will restrict ourselves to the case $M \geq N$. It is well known [13] that for large N the density of eigenvalues $\rho_N(\lambda) = \frac{1}{N} f(\frac{\lambda}{N})$ is given by the Marčenko-Pastur law:

$$f(x) = \frac{1}{2\pi x} \sqrt{(x - \zeta_-)(\zeta_+ - x)} \mathbb{1}_{x \in [\zeta_-, \zeta_+]} \quad (3)$$

with $\zeta_{\pm} = (1 \pm \sqrt{1 + \alpha})^2$, $\alpha = \frac{1-c}{c}$ and $c = \frac{N}{M}$. The indicator $\mathbb{1}_{x \in D}$ takes the value 1 if $x \in D$ and 0 otherwise. The points ζ_- and ζ_+ are usually called the hard edge and the soft edge of the distribution (3), respectively. The case $\alpha = 0$ (or $M = N$) corresponds to square matrices, and the case $\alpha = \mathcal{O}(1/N)$ (or $M - N = \mathcal{O}(1)$) is referred to here as "almost square matrices".

We will focus on the statistical fluctuations around the hard edge, which are captured by the probability distribution of the smallest eigenvalue $\rho_N^{(\min)}(\lambda)$, from which any other statistical property of the hard edge may be inferred. A classical result [14] shows that λ_{\min} converges

almost surely to ζ_- as $N \rightarrow \infty$. However, the concentration of the minimum around this value is generally unknown, and can be a challenging task.

The distribution of the smallest eigenvalue $\rho_N^{(\min)}(\lambda)$ (as well as for the maximal eigenvalue) has been formally expressed using Multivariate Hypergeometric Functions and Zonal Polynomials. This was first done long ago for real matrices ($\beta = 1$) in [15], then generalised to complex matrices ($\beta = 2$) in [16] and only recently to any β in [17]. These expressions are rigorous but often not easy to evaluate and manipulate (although a real breakthrough has occurred recently with a new algorithm that can calculate such quantities with complexity that is only linear in the size of the matrix [18], as well as available dedicated packages [19]).

Another line of explicit expressions involve a determinantal representation of the distribution of the smallest eigenvalue (as well as other order statistics) for finite N and M (see [20] and references therein). These expressions are not always easy to implement, and especially when $M - N$ is large.

More explicit expressions have been derived by Edelman in [21, 22] for $\beta = 1$ (for $\beta = 2$ as well, but only for square matrices). These expressions require the knowledge of some polynomials (different ones for any M and N). While manageable and useful for small matrices, this turns out to be impractical for large matrices, since no explicit formula exists for these polynomials. It is especially challenging to extract other relevant statistical properties around the hard edge like, for instance, the distribution of the typical (order $\mathcal{O}(N^p)$ with $0 < p \leq 1$) or large (order $\mathcal{O}(N)$) fluctuations around their mean values.

The typical fluctuations of the smallest eigenvalue have been recently studied rigorously in [23], where it is shown that the smallest eigenvalue λ_{\min} follows the Tracy-Widom (TW) distribution, that is, the typical fluctuations of λ_{\min} can be expressed as

$$\lambda_{\min} = \zeta_- N - \zeta_-^{2/3} c^{1/6} N^{1/3} \chi \quad (4)$$

where χ follows a TW distribution $g_\beta(\chi)$ [23]. This result was proven for the case $c < 1$ in the large N -limit, and so strictly speaking, it does not apply to either square or almost square matrices.

The knowledge about large deviations of the smallest eigenvalue is, as far as we are aware of, mainly unexplored. We would like here to correct the situation. To do so we will use the Coulomb gas approach [12]. Starting with Eq. (1) the cumulative probability of the minimum $P_N^{(\min)}(t) \equiv P(\lambda_{\min} \geq t)$ can be written as

$$P_N^{(\min)}(t) = \int_t^\infty d\lambda \rho_N^{(\min)}(\lambda) = \frac{Z(t)}{Z(0)} \quad (5)$$

with

$$Z(t) = \int_t^\infty \cdots \int_t^\infty d\lambda_1 \cdots d\lambda_N e^{-\beta F(\boldsymbol{\lambda})/2} \quad (6)$$

In this framework $Z(t)$ is understood as the partition function of a 2D Coulomb gas of charged particles restricted on a 1D line, in an external linear-log potential and a hard wall at t . The idea, as in [12], is to evaluate $Z(t)$ by using the saddle-point approximation. It is important to notice that the expression (5) is exact and that while the saddle-point method provides the exact density of eigenvalues for large N , it is only able to capture the large deviations to the right of the smallest eigenvalue from $\rho_N^{(\min)}(\lambda)$. This is not entirely surprising as $\rho_N(\lambda)$ is a collective quantity while $\rho_N^{(\min)}(\lambda)$ is not.

The calculation goes along similar lines as in [24, 25], so we shortly describe the needed steps. To apply the saddle-point method to (6), one first introduces the function $\varrho(\lambda) = \frac{1}{N} \sum_{i=1}^N \delta(\lambda - \lambda_i)$. This allows us to write the $Z(t)$ as a path integral over $\varrho(\lambda)$ and its Lagrange multiplier $\hat{\varrho}(\lambda)$. Minimising the corresponding functional with respect to $\hat{\varrho}(\lambda)$, allows to eliminate this multiplier and to unveil the meaning of $\varrho(\lambda)$ as a constrained spectral density. After rescaling the eigenvalues ($\lambda = xN$ and $\zeta = t/N$), $Z(t)$ takes the form

$$Z(t) = \int \{\mathcal{D}f\} e^{-\frac{\beta}{2} N^2 \mathcal{S}[f(x)]} \quad (7)$$

with the rescaled density $\varrho(\lambda) = \frac{1}{N} f(\frac{\lambda}{N})$, and the action

$$\begin{aligned} \mathcal{S}[f(x)] = & \int_\zeta^\infty dx f(x) x - \left(\alpha + \frac{\beta-2}{\beta N} \right) \int_\zeta^\infty dx f(x) \ln x \\ & - \int_\zeta^\infty \int_\zeta^\infty dx dx' f(x) f(x') \ln |x - x'| \\ & + \frac{2}{\beta N} \int_\zeta^\infty dx f(x) \ln f(x) + C_1 \left[\int_\zeta^\infty dx f(x) - 1 \right] \end{aligned} \quad (8)$$

As we will apply the saddle-point method we neglect the terms which are of order $\mathcal{O}(N^{-1})$ in the preceding expression for $\mathcal{S}[f(x)]$. Note however that while the first term in the subleading correction can be easily kept, the entropic term makes the evaluation of the saddle-point equation a challenging task. The next step is to look for the saddle point of the action. The saddle point equation is basically $\delta \mathcal{S}[f(x)] / \delta f(x) = 0$. It is useful to differentiate the saddle-point equation once with respect to x :

$$\frac{1}{2} - \frac{\alpha}{2x} = P \int_\zeta^\infty \frac{f_\star(x')}{x - x'} dx', \quad x \in [\zeta, \infty) \quad (9)$$

where $f_\star(x)$ is the value of $f(x)$ at the saddle-point. This is a Tricomi integral equation, and is solved as in [24], so we report the final result: For $\zeta \in [0, \zeta_-]$, $f_\star(x)$ is given by Eq. (3), while for $\zeta \in [\zeta_-, \infty)$,

$$f_\star(x) = \frac{\sqrt{U-x}}{2\pi\sqrt{x-\zeta}} \left(\frac{x - \alpha\sqrt{\zeta/U}}{x} \right) \mathbb{1}_{x \in [\zeta, U]} \quad (10)$$

with $U \equiv U(c, \zeta) = w^2(c, \zeta)$ and $w(c, \zeta)$ given by

$$w(c, \zeta) = \frac{2p}{3\rho^{1/3}} \cos \left(\frac{\theta + 2\pi}{3} \right), \quad (11)$$

with $p = -[\zeta + 2(\alpha + 2)]$, $q = 2\alpha\sqrt{\zeta}$, $\rho = \sqrt{-\frac{p^3}{27}}$, $\theta = \text{atan}\left(\frac{2\sqrt{B}}{q}\right)$ and $B = -\left(\frac{p^3}{27} + \frac{q^2}{4}\right)$. Using these results and after a long and tedious calculation we obtain the following results for the distribution: for $t \in [0, N\zeta_-]$ we have $P_N^{(\min)}(t) = 1$, and

$$P_N^{(\min)}(t) = e^{-\beta N^2 \Phi_+^{(\min)}\left(\frac{t - N\zeta_-}{N}\right)}, \quad t \in [N\zeta_-, \infty) \quad (12)$$

with the right rate function $\Phi_+^{(\min)}(x)$ being

$$\Phi_+^{(\min)}(x) = \frac{1}{2}[S(x + \zeta_-) - S(\zeta_-)], \quad x \in [0, \infty) \quad (13)$$

and the action $\mathcal{S}(\zeta) \equiv \mathcal{S}[f_\star(x)]$ given by

$$\begin{aligned} \mathcal{S}(\zeta) = & \frac{\zeta + U}{2} - \frac{\Delta^2}{32} - \ln\left(\frac{\Delta}{4}\right) + \frac{\alpha}{4}(\sqrt{U} - \sqrt{\zeta})^2 \\ & + \frac{\alpha^2}{4} \ln(\zeta U) - \alpha(\alpha + 2) \ln\left[\frac{\sqrt{U} + \sqrt{\zeta}}{2}\right], \end{aligned} \quad (14)$$

where, $\Delta = U(c, \zeta) - \zeta$. As mentioned above, the saddle-point approximation is only able to capture the large fluctuations to the right of λ_{\min} as it has a hard wall on the left side (12). Fortunately, the authors of [26] came up with a beautiful physical argument to overcome this shortcoming and to estimate in their case the large deviations from the maximal eigenvalue. Applied to the fluctuation to the left of λ_{\min} this method yields

$$P_N^{(\min)}(t) \sim e^{-\beta N \Phi_-^{(\min)}\left(\frac{N\zeta_- - t}{N}\right)}, \quad t \in [0, N\zeta_-] \quad (15)$$

with the left rate function $\Phi_-^{(\min)}(x)$ given by:

$$\begin{aligned} \Phi_-^{(\min)}(x) = & -\frac{\alpha}{2} \ln\left(1 - \frac{x}{\zeta_-}\right) - \frac{1}{2}\sqrt{x}\sqrt{x + \Delta_-} \\ & + 2 \ln\left(\frac{\sqrt{x + \Delta_-} - \sqrt{x}}{\sqrt{\Delta_-}}\right) + \alpha \ln\left(1 + 2\sqrt{\frac{x}{\zeta_-}} \frac{\sqrt{x + \Delta_-} - \sqrt{x}}{\Delta_-}\right) \end{aligned} \quad (16)$$

with $x \in [0, \zeta_-]$ and $\Delta_- = \zeta_+ - \zeta_- = 4\sqrt{\alpha + 1}$.

We now check that the smallest large fluctuations predicted by our results match the largest typical fluctuations given by the TW distribution [23]. Expanding the rate functions $\Phi_\pm^{(\min)}(x)$ around $x = 0$ gives:

$$\Phi_-^{(\min)}(x) \underset{x \rightarrow 0}{\sim} \frac{2}{3\zeta_- c^{1/4}} x^{3/2} \quad (17)$$

$$\Phi_+^{(\min)}(x) \underset{x \rightarrow 0}{\sim} \frac{1}{24\zeta_-^2 \sqrt{c}} x^3 \quad (18)$$

which yields the following expression of $P_N^{(\min)}(t)$ for the smallest large fluctuations of λ_{\min} :

$$P_N^{(\min)}(t) \sim \begin{cases} \exp\left(-\frac{2\beta}{3}\chi^{3/2}(t)\right), & t \in [0, N\zeta_-] \\ \exp\left(-\frac{\beta}{24}|\chi(t)|^3\right), & t \in [N\zeta_-, \infty) \end{cases}$$

with $\chi(t) \equiv -(N\zeta_- - t)/(N^{1/3}\zeta_-^{2/3}c^{1/6})$. This result obviously agrees with TW distribution for large $|\chi|$ [23, 27].

For larger atypical fluctuations, we have the following asymptotic behaviours of the rate functions:

$$\Phi_-^{(\min)}(x) \underset{x \rightarrow \zeta_-}{\sim} \left[-\frac{\zeta_-}{2} - \sqrt{\zeta_-} + \frac{\ln(c)}{2c} + \alpha \ln \alpha\right] - \frac{\alpha}{2} \ln(\zeta_- - x) + \mathcal{O}(\zeta_- - x), \quad (19)$$

$$\Phi_+^{(\min)}(x) \underset{x \rightarrow \infty}{\sim} \frac{1}{2}x - \frac{\alpha}{2} \ln x + \left[\frac{\zeta_-}{2} + \frac{3}{4} - \frac{S(\zeta_-)}{2}\right]. \quad (20)$$

Interestingly, these results can be compared with the exact asymptotic behaviours predicted in [22] for the case $\beta = 1$ and large N . It turns out that for $\Phi_-^{(\min)}(x)$ the leading logarithmic behaviour agrees perfectly with [22], while the constant term cannot be rigorously compared, since it is not available for any N in the exact treatment [22] (although we know that a constant term exists). For $\Phi_+^{(\min)}(x)$ we again see perfect agreement with the linear and logarithmic terms, but the constant deviates from the exact one given in [22].

A comparison of these results, as well as with results of simulations and the TW distribution [28] are summarized in Fig. 1. Note that exact results are only available for $\beta = 1$ [22], and so a comparison to exact results for ensembles other than the real Wishart case is not possible.

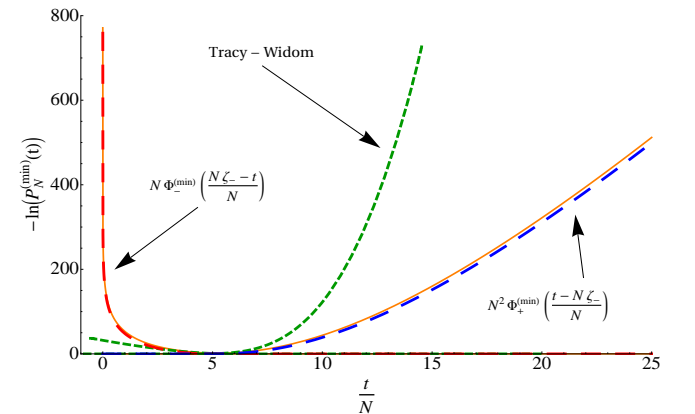


FIG. 1. (color online). Results for the smallest eigenvalue distribution $-\ln P_N^{(\min)}(t)$ vs. the scaled variable t/N . Here, $N = 11$ and $M = 110$ ($c = 1/10$), Wishart matrices are real ($\beta = 1$). The large deviation functions (dashed lines) compare very well with the exact result of [22] (solid line), while TW (dotted line) describes only small fluctuations.

We now pay special attention to the case of almost square matrices, namely when $M = N + a$ with a an integer of order unity (i.e. $\alpha = a/N$, so α is of order $1/N$). Here it is useful to use the scaling $z = Nt = N^2\zeta$ to unravel non-trivial results. Using the Coulomb gas approach we find that in this special case, the PDF of λ_{\min} does not simply approach a delta function, in the large N limit, as in [14]. Instead, the cumulative distribution of z has a N -independent limiting shape as shown in Fig. 2 for $\beta = 1$, and whose large fluctuations are described by the

Coulomb gas prediction:

$$P_N^{(\min)}(z) \sim \begin{cases} \exp\left(-\beta a \Psi_-^{(\min)}\left(\frac{4z}{a^2}\right)\right), & z \in [0, a^2/4] \\ \exp\left(-\beta a^2 \Psi_+^{(\min)}\left(\frac{4z}{a^2}\right)\right), & z \in [a^2/4, \infty) \end{cases} \quad (21)$$

with $\Psi_+^{(\min)}(x) = (x - 4\sqrt{x} + 3 + \ln x)/8$ and $\Psi_-^{(\min)}(x) = \ln \frac{1+\sqrt{1-x}}{\sqrt{x}} - \sqrt{1-x}$. The functions $\Psi_{\pm}^{(\min)}(x)$ are universal, in the sense that they are N, β and a -independent. Note that in this particular regime one needs to keep the first $1/N$ term appearing (8) as it is of the same order as $\alpha = a/N$. This can be accounted for easily by replacing in the preceding expressions $a \rightarrow a(\beta) = a + (\beta - 2)/\beta$, as long as $a(\beta)$ is non-negative. To our knowledge, this is the first time that this case is discussed and shown to be universal (even numerically), and obviously no general explicit predictions for its shape have been proposed, apart of the particular case of $a(\beta) = 0$ for which Eq. (21) yields $P_N^{(\min)}(z) = e^{-\beta z/2}$ in agreement with the exact result of [22] (for $a = \beta = 1$) and [21] (for $a = 0, \beta = 2$). Note also that the results for TW distribution reported in [23] do not formally apply to almost square matrices. In Fig. 2 we have compared our finding with Edelman's exact result for $\beta = 1, N = 200$ and $M = 205$.

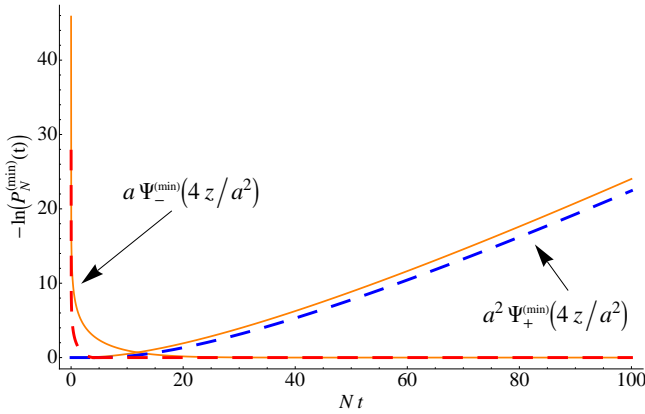


FIG. 2. (color online). Results for the smallest eigenvalue distribution $-\ln P_N^{(\min)}(t)$ vs. the scaled variable tN . Here, $N = 200$ and $a = 5$, Wishart matrices are real ($\beta = 1$). The large deviation functions derived here (dashed lines) compare well with the exact result of [22] (solid line).

It is important to point out that while the results in [22] are exact, they become difficult to evaluate for large values of N and a daunting task to extract exact results about large fluctuations for large matrices. In contrast, the result (21) provide information of large fluctuations for any value of β and a , when $a(\beta) \geq 0$.

To summarise, in this work we study the large deviation functions of the smallest eigenvalue of the Wishart-Laguerre Ensemble using the Coulomb gas approach. We

obtain explicit expressions for both the right and left rate functions for general α . We also highlight the existence of a special regime for almost square matrices, where an interesting limit distribution exists, described by universal rate functions $\Psi_{\pm}^{(\min)}(x)$. We were able to provide predictions for $a(\beta) \geq 0$, which leaves the question of $a(\beta) < 0$ open for further research. Another interesting open question is regarding the typical fluctuations for the smallest eigenvalue for square and almost square matrices, which are not captured by the TW distribution.

-
- [1] E. P. Wigner, Proc. Cambridge Philos. Soc., **47**, 790 (1951).
 - [2] J. Wishart, Biometrika, **20A**, 32 (1928).
 - [3] G. Akemann and E. Kanzieper, Phys. Rev. Lett., **85**, 1174 (2000).
 - [4] T. Rogers, I. Pérez Castillo, R. Kühn, and K. Takeda, Phys. Rev. E, **78** (2008).
 - [5] A. M. Tulino and S. Verdú, *Random matrix theory and wireless communications* (now Publishers Inc., Hanover, 2004).
 - [6] M. B. Eisen, P. T. Spellman, P. O. Brown, and D. Botstein, Proc. Natl. Acad. Sci. USA, **95**, 14863 (1998).
 - [7] C. N. et al, Arxiv:1006.4091 (2010).
 - [8] A. E. Litvak, A. Pajor, M. Rudelson, and N. Tomczak-Jaegermann, Adv. Math., **195**, 491 (2005).
 - [9] M. Rudelson, Ann. Math., **168**, 575 (2008).
 - [10] E. J. Candes and T. Tao, IEEE Trans. Inform. Theory, **52**, 5406 (2006).
 - [11] R. J. Muirhead, *Aspects of multivariate statistical theory* (Wiley, 1982).
 - [12] D. S. Dean and S. N. Majumdar, Phys. Rev. Lett., **97** (2006); Phys. Rev. E, **77** (2008).
 - [13] V. A. Marcenko and L. A. Pastur, Math. USSR-Sb, **1**, 457 (1967).
 - [14] J. W. Silverstein, Ann. Probab., **13**, 1364 (1985).
 - [15] P. R. Krishnaiah and T. C. Chang, Ann. Inst. Statist. Math., **23**, 293 (1971).
 - [16] T. Ratnarajah, R. Vaillancourt, and M. Alvo, SIAM J. Matrix Anal. Appl., **26**, 441 (2004).
 - [17] I. Dumitriu and P. Koev, SIAM J. Matrix Anal. Appl., **30**, 1 (2008).
 - [18] P. Koev and A. Edelman, Math. Comp., **75**, 833 (2006).
 - [19] I. Dumitriu, A. Edelman, and G. Shuman, J. Symbolic Comput., **42**, 587 (2007).
 - [20] G. Akemann and P. Vivo, J. Stat. Mech., P09002.
 - [21] A. Edelman, SIAM J. Matrix Anal. Appl., **9**, 543 (1988).
 - [22] A. Edelman, Linear Algebra Appl., **159**, 55 (1991).
 - [23] O. N. Feldheim and S. Sodin, Geom. Funct. Anal., **20**, 1 (2010).
 - [24] P. Vivo, S. N. Majumdar, and O. Bohigas, J. Phys. A, **40**, 4317 (2007).
 - [25] Y. Chen and S. M. Manning, J. Phys. A, **27**, 3615 (1994).
 - [26] S. N. Majumdar and M. Vergassola, Phys. Rev. Lett., **102**, 060601 (2009).
 - [27] C. A. Tracy and H. Widom, Comm. Math. Phys., **159**, 151 (1994); **177**, 727 (1996).
 - [28] We have used the table provided by Prähofer and Spohn at <http://www-m5.ma.tum.de/KPZ/>.